Sur la longueur des mots de rang donné d'un automate fini

Jean-Eric Pin

INFORMATIQUE THÉORIQUE. *Sur la longueur des mots de rang donné d'un automate fini.* Note* de **Jean-Éric Pin**, présentée par M. André Lichnerowicz.

$\mathcal{A}$ étant un automate fini à n états, nous montrons que s'il existe un mot de rang inférieur ou égal à $(n-k)$ dans $\mathcal{A}$, il en existe en particulier un de longueur inférieure ou égale à $P(k)$, où P est un polynôme de degré 4.

Let $\mathcal{A}$ be a finite automaton with n states. We prove that if there exists a word of rank $\leq (n-k)$ in $\mathcal{A}$, then there exists such a word with length $\leq P(k)$, where P is a polynomial of degree 4.

Soit $\mathcal{A} = (Q, X)$ un automate fini, où Q est l'ensemble des états et X l'alphabet. On notera qm l'action d'un mot m de X^* sur l'état s et si S est un sous-ensemble de Q, Sm l'ensemble $\{qm \mid q \in S\}$. $|B|$ désignera le cardinal de l'ensemble B, et si n est un mot de X^*, $|m|$ désignera la longueur de m, le contexte évitant toute confusion entre ces deux notations. Le rang de m dans $\mathcal{A}$ est $r_{\mathcal{A}}(m) = |Qm|$. Soit $\Pi_{n,k}$ l'ensemble des automates à n états possédant un mot de rang inférieur ou égal à k. On pose

$$A(n,k) = \sup_{\mathcal{A} \in \Pi_{n,k}} \inf\{|m| \mid r_{\mathcal{A}}(m) \leq k\}$$

Différents auteurs, comme J. Černỳ, P.H. Starke — ont donné diverses bornes supérieures pour $A(n, 1)$, dont la meilleure est

$$A(n,1) \leq \frac{1}{3}n^3 - \frac{3}{2}n^2 + \frac{25}{6}n - 4.$$

Mais les mêmes méthodes appliquées à $A(n, n-k)$ donnent une majoration de $A(n, n-k)$ dépendant de n et de k (sauf si $k = 0, 1, 2$). Nous démontrons que $A(n, n-k)$ peut être majoré par un polynôme dépendant uniquement de k.

Theorem 1 $A(n, n-k) \leq \frac{k(k-1)(k^2+k+6)}{12}$.

Ce théorème repose sur le résultat suivant, qui présente un intérêt par lui-même :

Theorem 2 *Soit $\mathcal{A} = (Q, X)$ un automate à n états ; Q' une partie à k éléments de Q. S'il existe un mot m tel que $|Q'm|$ soit inférieur ou égal à $k-1$, alors il existe un tel mot de longueur inférieure ou égale à*

$$\frac{1}{3}(n-k)^3 + \frac{1}{2}(n-k)^2 + \frac{7}{6}(n-k) + 1$$

Preuve du théorème 2. — Soit m un mot de longueur minimale tel que $|Q'm| \leq k-1$. Posons $m = x_1 \cdots x_p$, $Q_1 = Q'$, $Q_2 = Q'x_1$, …, $Q_p = Q_1 x_1 \cdots x_{p-1}$. On a

$$|Q_1| = |Q_2| = \cdots = |Q_p| = k$$

sinon m ne serait pas de longueur minimale. En outre puisque $|Q_p x_p| \leq k-1$, Q_p contient une paire $\{a_p, b_p\}$ telle que $a_p x_p = b_p x_p$. Il existe donc des paires $\{a_i, b_i\}$ incluses dans Q_i telles que $a_i x_i = a_{i+1}$, $b_i x_i = b_{i+1}$ pour $i = 1, \ldots, p-1$.

Compte tenu de la définition de m, on a la propriété suivante :

$$(*) \qquad \text{Si } j < i,\ \{a_i, b_i\} \not\subset Q_j$$

*Séance du 28 mars 1977.

Sinon on aurait $\{a_i, b_i\} \subset Q'x_1 \cdots x_{j-1}$. Or

$$a_i x_i \cdots x_p = b_i x_i \cdots x_p, \text{ donc } |Q'x_1 \cdots x_{j-1} x_i \cdots x_p| \leq k-1$$

mais $x_1 \cdots x_{j-1} x_i \cdots x_p$ est plus court que m d'où une contradiction. Il reste à majorer $p = |m|$. Pour cette majoration, il est plus commode de raisonner sur les complémentaires des Q_i. Posons donc $L_i = Q \setminus Q_i$ et $t = n - k$. D'après $(*)$ les $(L_i)_{i=1}^p$ vérifient

$$(**) \qquad \begin{cases} \text{Pour tout } i \in \{1, \ldots, p\}, \text{ il existe } a_i \text{ et } b_i \text{ dans } Q \text{ tels que} \\ a_i \neq b_i,\ \{a_i, b_i\} \cap L_i = \emptyset \text{ et, pour tout } j < i,\ \{a_i, b_i\} \cap L_j \neq \emptyset. \end{cases}$$

Lemme 3 *Soit Q un ensemble, $(L_i)_{i=1}^{p_t}$ une suite de sous-ensembles à t éléments de Q vérifiant les conditions $(**)$. Alors*

$$p_t \leq \frac{1}{3}t^3 + \frac{1}{2}t^2 + \frac{7}{6}t + 1.$$

La démonstration se fait par récurrence sur l'ensemble des couples $\{(t, q) \mid q \leq t\}$ ordonné suivant l'ordre lexicographique. On montre par récurrence les deux propriétés suivantes pour $(t, q) \geq (1, 0)$:

(1) si les L_i sont de cardinal t et si

$$p_t \geq k_q = 1 + q + \sum_{j=1}^{q-1} j^2, \text{ alors } |\bigcap_{i=1}^{k_q} L_i| \leq t - q;$$

(2) si les L_i sont de cardinal $(t-1)$, alors $p_{t-1} \leq t + \sum_{1 \leq j \leq t-1} j^2$.

$(t, q) = (1, 0)$. — Les deux conditions sont évidentes. On suppose le résultat vrai jusqu'au rang (t, q) et on cherche à le démontrer pour son successeur.

Passage de (t, q) à $(t, q+1)$ (si $q \leq t-1$). — Le (2), indépendant de q, résulte de l'hypothèse de récurrence. Montrons le (1) par l'absurde : supposons $p_t \geq k_q$ et $|\bigcap_{1 \leq i \leq k_{q+1}} L_i| = t - q$ et posons $\bigcap_{1 \leq i \leq k_{q+1}} L_i = K$. La suite des $(L_i \setminus K)$ $(1 \leq i \leq k_{q+1})$ est une suite de sous-ensembles à q éléments $(q \leq t-1)$ qui vérifie $(**)$. En effet, puisque $\{a_i, b_i\} \cap L_i = \emptyset$, $\{a_i, b_i\} \cap (L_i \setminus K) = \emptyset$ et $\{a_i, b_i\} \cap K = \emptyset$. Donc pour tout $j < i$, $\{a_i, b_i\} \cap (L_j \setminus K) = \{a_i, b_i\} \cap L_j \neq \emptyset$. D'après l'hypothèse de récurrence (2), la suite des $(L_i \setminus K)$ (i variant de 1 à k_{q+1}) possède au plus $(1 + q + \sum_{1 \leq j \leq q} j^2)$ termes ; or $k_{q+1} = 2 + q + \sum_{1 \leq j \leq q} j^2$, d'où une contradiction.

Passage de (t, t) à $(t+1, 0)$. — Dans ce cas, (1) est évident. Si $p_t < k_t$, (2) est démontré. Sinon, on peut appliquer l'hypothèse de récurrence (1) au rang (t, t) : les L_i étant de cardinal t, $\bigcap_{1 \leq i \leq k_t} L_i = \emptyset$. Les L_i d'indice $> k_t$ vérifient en particulier

$$\exists a_i\ \exists b_i\ \{a_i, b_i\} \cap L_i = \emptyset \text{ et } \{a_i, b_i\} \cap L_{k_t} \neq \emptyset.$$

Donc l'un des éléments de la paire, disons a_i, est pris parmi les t éléments de L_{k_t}. En outre, puisque $\bigcap_{1 \leq i \leq k_t} L_i = \emptyset$, il existe un L_j (avec $j < k_t$) tel que $a_i \notin L_j$; comme $\{a_i, b_i\} \cap L_j$ est non vide, on a $b_i \in L_j$; b_i est donc pris parmi les t éléments de L_i. Il y a donc au plus t^2 paires vérifiant les conditions ci-dessus, et donc au plus t^2 indices strictement supérieurs à k_t. Donc

$$p_t \leq k_t + t^2 \leq 1 + t + \sum_{1 \leq j \leq q} j^2 \leq \frac{1}{3}t^3 + \frac{1}{2}t^2 + \frac{7}{6}t + 1,$$

ce qui achève les démonstrations du lemme et du théorème 2.

Preuve du théorème 1. — Soit m un mot de rang inférieur ou égal à $n-k$. Si Q' est une partie quelconque de Q, $|Q'm| \leq n-k$ et d'après le théorème 2, il existe un mot m_1 de longueur inférieure ou égale à $P(0)$ (où P est le polynôme $\frac{1}{3}X^3+\frac{1}{2}X^2+\frac{7}{6}X+1$) et de rang inférieur ou égal à $n-1$. Donc $|Qm_1| \leq n-1$ et il existe un mot m_2 de longueur inférieure ou égale à $P(1)$ tel que $|Qm_1m_2| \leq n-2$. On construit ainsi une séquence de mots $m_1, \ldots, m_k$ tels que $|m_i| \leq P(i)$ et $|Qm_1m_2 \cdots m_k| \leq n-k$. Donc $m = m_1m_2 \cdots m_k$ est de rang inférieur ou égal à $n-k$ et

$$|m| \leq \sum_{0 \leq i \leq k-1} P(i) \leq \frac{k(k+1)(k^2+k+6)}{12}.$$

Références

[1] J. Černý, Poznámka k. homogénnym experimentom s konecnými automatmi, *Mat. fyz. čas SAV* **14** (1964), 208–215.

[2] J. Černý, Communication, in *Bratislava Conference on Cybernetics*, 1969.

[3] J. Černý, A. Pirická and B. Rosenauerova, On directable automata, *Kybernetica* **7** (1971), 289–298.

[4] P. H. Starke, Eine Bemerkung über homogene Experimente., *Elektr. Informationverarbeitung und Kyb.* **2** (1966), 257–259.

[5] P. H. Starke, *Abstrakte Automaten*, V.E.B. Deutscher Verlag der Wissenschaften, Berlin, 1969.